



Kasimir : gestion de connaissances décisionnelles en cancérologie

Mathieu d'Aquin, Sébastien Brachais, Jean Lieber, Amedeo Napoli

► To cite this version:

Mathieu d'Aquin, Sébastien Brachais, Jean Lieber, Amedeo Napoli. Kasimir : gestion de connaissances décisionnelles en cancérologie. Modélisation et pilotage des systèmes de Connaissances et de Compétences dans les Entreprises Industrielles - C2EI'04, 2004, Nancy, France, 6 p. inria-00099935

HAL Id: inria-00099935

<https://inria.hal.science/inria-00099935>

Submitted on 26 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

KASIMIR : GESTION DE CONNAISSANCES DÉCISIONNELLES EN CANCÉROLOGIE

d'Aquin Mathieu, Brachais Sébastien, Lieber Jean, Napoli Amedeo

LORIA, UMR 7503, CNRS, INRIA, Universités de Nancy,
Loria BP 239, 54 506 Vandœuvre-lès-Nancy
Mél.: {daquin, brachais, lieber, napoli}@loria.fr

RÉSUMÉ : Cet article présente le projet de recherche KASIMIR pour la gestion des connaissances décisionnelles en cancérologie, sous certains de ses aspects informatiques. Ces connaissances se présentent sous la forme de référentiels (ou protocoles de décision) semblables à des arbres de décision et représentés, au sein du système KASIMIR, dans un formalisme par objets. Un moteur d'inférences fondé sur ce formalisme et sur la classification hiérarchique est associé à un éditeur de connaissances. Cette association permet une édition et une maintenance de référentiels, une détection de certaines erreurs d'édition et une comparaison entre versions successives du référentiel. De cette façon, la gestion des référentiels tire pleinement avantage des outils de représentation des connaissances et de raisonnement.

MOTS-CLÉS : représentation de connaissances, aide à la décision, référentiels médicaux

1 INTRODUCTION

Cet article présente KASIMIR, un projet multidisciplinaire de gestion de connaissances décisionnelles en cancérologie sous ses aspects informatiques. Le sous-domaine de la cancérologie le plus étudié dans KASIMIR est le traitement du cancer du sein qui s'appuie sur un *référentiel* (qu'on peut assimiler à une base de connaissances pour l'aide à la décision). L'objectif de l'article est de montrer comment la représentation des connaissances et le raisonnement sont utiles pour la gestion des connaissances et, en particulier, pour l'édition et la maintenance des référentiels.

Un survol du projet KASIMIR est présenté à la section 2. La section 3 présente le module de raisonnement du système KASIMIR et son formalisme de représentation des connaissances. La section 4 montre l'apport de ce raisonnement artificiel pour assister l'édition et la maintenance de connaissances. La section 5 présente les recherches actuelles sur l'adaptation du référentiel aux cas pour lesquels sont application directe ne donne pas satisfaction.

2 LE PROJET KASIMIR

L'objectif du système KASIMIR est d'assister la prise de décision et la gestion des connaissances pour le traitement du cancer (voir (Lieber *et al.*, 2002), pour plus de détail). Ce système est actuellement en développement dans le cadre d'un projet multidisciplinaire dans lequel participent des chercheurs en informatique (auteurs de cet article) et en psycho-ergonomie (du Laboratoire d'ergonomie du CNAM, à Paris), des experts en cancérologie (du *Centre Alexis Vautrin* à Vandœuvre-lès-Nancy) et ONCOLOR, une association regroupant les médecins lorrains impliqués dans la cancérologie.

Certains référentiels (ou protocoles de décision) sont disponibles pour résoudre des problèmes de décision en cancérologie (voir www.oncolor.org). Ces référentiels sont construits selon les principes de la médecine factuelle (*evidence-based medicine* (Evidence-based medicine working-group, 1992; Sauvagnac *et al.*, 1999)). Plusieurs référentiels ont été implantés dans KASIMIR (notamment pour la surveillance post-thérapeutique des cancers du sein et de la prostate) et l'acquisition et l'implantation d'autres référentiels sont en cours. Un de ces référentiels est consacré au traitement du cancer du sein non métastatique et, dans la suite du papier, il sera appelé « le » référentiel. Le système KASIMIR propose des traitements sur la base d'un référentiel codé informatiquement. Son implantation s'appuie sur un formalisme de représentation des connaissances par objets et sur la classification hiérarchique.

Pour la plupart des cas (environ 70%), une application directe du référentiel est suffisante : il donne un traitement qui peut être directement appliqué. Un cas des 30% de cas restant est dit « hors référentiel ». Cela signifie soit que le référentiel ne propose pas de traitement pour ce cas (p. ex., quand le patient est un homme : le référentiel a été conçu pour les femmes, mais certains hommes sont atteints du cancer du sein), soit que le traitement proposé par le référentiel soulève certaines difficultés (contre-indica-

tions, impossibilité d'appliquer complètement un traitement, etc.). Pour ces cas hors référentiel, les cancérologues tentent d'adapter le référentiel durant des réunions des « Comités de concertation pluridisciplinaires » (CCP) qui réunissent des experts des spécialités liées à la cancérologie du sein (la chimiothérapie, la chirurgie, la gynécologie, la radiothérapie, etc.). L'adaptation est actuellement étudiée sous les points de vue de la psycho-ergonomie et de l'informatique. La conception et le développement de fonctionnalités d'adaptation dans le système est en cours et se fait selon les principes du raisonnement à partir de cas (RÀPC (Riesbeck et Schank, 1989)).

Les adaptations peuvent être utilisées afin de proposer des évolutions du référentiel grâce à sa confrontation avec des cas réels. L'idée est alors de suggérer des évolutions du référentiel fondé sur les adaptations fréquemment réalisées en CCP. Ceci a été étudié par les psycho-ergonomes du projet (Sauvagnac, 2000). L'implantation d'un système informatique qui puisse faire de telles suggestions est une perspective à long terme.

Le système KASIMIR est prévu pour être utilisé par des médecins dans leurs pratiques quotidiennes. Une validation expérimentale a d'ores et déjà été réalisée (Rios *et al.*, 2003). Cette étude avait pour but de voir comment l'utilisation de ce système permettait d'améliorer la qualité des soins. 30 médecins avaient à proposer des traitements pour un échantillon de cas à l'aide du référentiel sous sa forme papier (le référentiel dessiné sous la forme d'un arbre de décision associé à des explications textuelles) ou à l'aide de KASIMIR (avec le référentiel sous forme d'une base de connaissances). Une amélioration statistiquement significative de la conformité aux standards médicaux quand KASIMIR est utilisé plutôt que le référentiel papier a été montrée (Rios *et al.*, 2003). Selon le paradigme de la médecine factuelle (Evidence-based medicine working-group, 1992), l'utilisation de KASIMIR devrait donc améliorer la qualité des soins.

3 RAISONNEMENT ET REPRÉSENTATION DES CONNAISSANCES

Le principe du développement du système KASIMIR est la genericité, au sens où son adaptation à différents référentiels est voulue la plus simple possible. Pour ce faire, la base de connaissances et les spécifications de l'interface utilisateur sont décrits par des fichiers XML. Tout changement se fait à travers la modification de ces fichiers. La figure 1 montre l'interface pour envoyer des requêtes au moteur d'inférences et

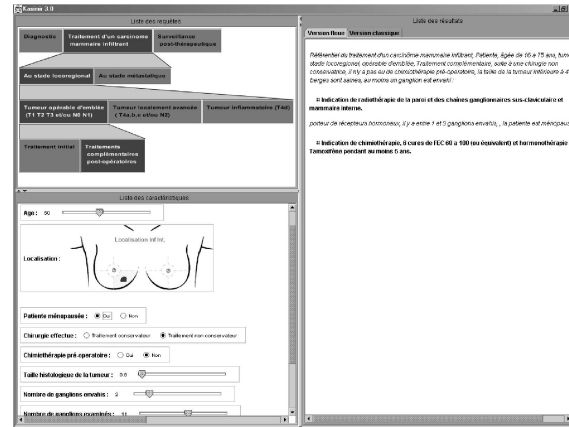


FIG. 1 – L'interface utilisateur de KASIMIR. Le panneau en haut à gauche est utilisé pour choisir un sous-référentiel du référentiel (p. ex., traitement initial, traitement après chirurgie, etc.). Le panneau en bas à gauche est utilisé pour saisir des caractéristiques du patient (âge, etc.) et de sa tumeur (taille, etc.). Le panneau de droite permet l'affichage des propositions de traitements associées à des explications ; il est mis à jour à chaque action de l'utilisateur sur un des panneaux de gauche.

afficher ses conclusions.

La représentation des connaissances est au cœur de la gestion des connaissances du système KASIMIR. Elle s'appuie sur un formalisme de représentation des connaissances par objets (Napoli *et al.*, 1994) proche d'une logique de descriptions (Baader *et al.*, 2003). L'unité de représentation de base est le *concept*, qui représente un ensemble d'individus partageant des *attributs*. L'ensemble des individus représenté par un concept C est appelée l'*extension* de C alors que l'ensemble de ses attributs est appelé l'*intension* de C. Un individu de l'extension de C est appelé *instance* de C. Un attribut a un *domaine*, qui est le concept C auquel il est attaché, et un *co-domaine*, qui détermine le type des valeurs admissibles de cet attribut pour les individus de l'extension de C. Ce co-domaine peut être de type numérique ou être un autre concept. Dans ce dernier cas, un attribut définit une *relation* entre deux concepts : le domaine et le co-domaine de l'attribut.

Deux types de concepts sont considérés. Les *concepts primitifs* sont considérés comme les atomes du système de représentation. Ils sont utilisés comme « briques de construction » pour les concepts définis. De plus, l'intension d'un concept primitif est vide (il n'a aucun attribut) alors que l'intension d'un concept défini est l'ensemble des attributs et peut être vu comme un ensemble de conditions nécessaires et suffi-

$$\begin{array}{l|l} \text{booléen-qcq} \sqsubseteq \top & \text{vrai} \sqsubseteq \text{booléen-qcq} \\ \text{grade-qcq} \sqsubseteq \top & \text{faux} \sqsubseteq \text{booléen-qcq} \\ \text{traitement-qcq} \sqsubseteq \top & \text{gI} \sqsubseteq \text{grade-qcq} \\ \text{chimio-niveau-1} \sqsubseteq \text{traitement-qcq} & \\ \text{pas-de-chimio} \sqsubseteq \text{traitement-qcq} & \\ \text{FN_RH}_+ \text{G}_I = (\text{sexe} : \text{féminin} \sqcap \text{N} : \text{faux} & \\ \quad \sqcap \text{RH} : \text{vrai} \sqcap \text{tumeur} : (\text{grade} : \text{gI})) & \\ \text{A}_{<35} = \text{FN_RH}_+ \text{G}_I \sqcap (\text{âge} : [0,35]) & \end{array}$$

$A_{>35} = FN_RH + G_I \sqcap (\text{âge} :]35, +\infty[)$
 $T_{\geq 1} = A_{\leq 35} \sqcap (\text{tumeur} : (\text{taille} : [1; +\infty[))$
 $A_{\geq 70} = A_{>35} \sqcap (\text{âge} : [70, +\infty[)$

Les règles $R = (pb \longrightarrow \text{Sol}(pb))$ de ce référentiel sont :

$R_1 = (T_{\geq 1} \longrightarrow \text{chimio-niveau-1})$
 $R_2 = (A_{\geq 70} \longrightarrow \text{pas-de-chimio})$

La classification du problème *cible* dans la hiérarchie des concepts considérée comme une hiérarchie de problèmes retourne l'ensemble des problèmes qui subsument *cible*. Dès qu'un de ces subsumants a une solution associée, cette solution peut être réutilisée dans le contexte de *cible*. Par exemple, considérons la patiente décrite par l'axiome suivant :

cible = (sexe : féminin \sqcap N : faux \sqcap RH : vrai
 \sqcap âge : [75,75] \sqcap tumeur : (grade : gI))

La classification de *cible* dans la hiérarchie des problèmes montre qu'elle est subsumée par les concepts suivants : $FN_RH + G_I$, $A_{>35}$ et $A_{\geq 70}$. La solution *pas-de-chimio* est associée à ce dernier par la règle R_2 . Par conséquent, le système KASIMIR recommande que cette patiente n'ait pas de chimiothérapie.

Le formalisme de représentation présenté ci-dessus est assez simple (comparé aux logiques de descriptions telles que *ALC* (Baader *et al.*, 2003)) mais il est suffisant pour les référentiels représentés jusqu'à présent dans KASIMIR. Un avantage de cette simplicité est la faible complexité en temps de ces inférences : dans le pire cas, la classification est en $O(n_a n_c)$ où n_a et n_c sont les nombres d'attributs et de concepts (quand la hiérarchie est un arbre équilibré, cette complexité est en $O(n_a \log n_c)$). En pratique, même pour le référentiel le plus complexe actuellement représenté dans le système KASIMIR (le référentiel pour le traitement du cancer du sein non métastatique qui est représenté par environ 1200 concepts et 50 attributs), le résultat de ce processus de raisonnement est donné de façon immédiate sur un ordinateur personnel actuel.

4 ÉDITION ET MAINTENANCE DE CONNAISSANCES

Représenter un référentiel dans KASIMIR peut être fastidieux si celui-ci est important, ce qui a entraîné le besoin d'utiliser un éditeur de connaissances pour KASIMIR. Le système PROTÉGÉ (Noy *et al.*, 2000) a été choisi dans ce but, en particulier parce que beaucoup d'outils utiles et disponibles sont intégrés dans son architecture. PROTÉGÉ a été adapté en un outil d'édition pour KASIMIR et a été connecté à son moteur d'inférences.

PROTÉGÉ et KASIMIR ont été connectés. La première

étape de cette connexion a été d'intégrer le modèle de représentation des connaissances de KASIMIR dans PROTÉGÉ. Alors, un référentiel peut être édité (créé ou mis à jour) et exporté sous une forme de fichiers XML exploitables par KASIMIR. De plus, PROTÉGÉ a été connecté au moteur d'inférences de KASIMIR ce qui permet en particulier de détecter des erreurs durant les sessions d'édition des connaissances. Par exemple, il peut arriver que deux problèmes pb_1 et pb_2 soient édités avec deux définitions équivalentes (i.e., ils représentent le même ensemble d'individus : $pb_1 \sqsupseteq pb_2$ et $pb_2 \sqsupseteq pb_1$). Dans une telle situation, le moteur d'inférences détecte cette équivalence et l'utilisateur est alerté. Un autre exemple est quand la hiérarchie *déclarée* dans PROTÉGÉ n'est pas isomorphe à la hiérarchie *calculée* dans KASIMIR : cette non-correspondance signifie souvent qu'il y a eu erreur d'édition. Ces alertes ont montré leur utilité pratique en permettant de détecter beaucoup d'erreurs d'édition.

Deux modules de visualisation ont été intégrés dans PROTÉGÉ pour permettre l'affichage de la hiérarchie des problèmes calculée par KASIMIR, lors de l'édition : PALÉTUVIER et HYPERTREE (voir figure 3). L'utilisation combinée de ces deux modules ainsi que de la visualisation des arbres de PROTÉGÉ permet une visualisation des hiérarchies facilitant la navigation, une vue globale et une vue focalisée sur un nœud particulier.

Enfin, un module pour comparer deux versions d'un même référentiel, KILT, a été développé : il permet de différencier les problèmes de l'ancien référentiel et les problèmes du nouveau, et de visualiser cette différence en s'appuyant sur PALÉTUVIER.

Pour plus de détail sur l'édition et la maintenance de connaissances pour KASIMIR, voir (d'Aquin *et al.*, 2003)

5 ADAPTATION DANS KASIMIR

Comme cela a été mentionné à la section 2, durant les réunions du CCP, le référentiel est *adapté* pour les cas hors référentiels. Afin d'assister les experts du CCP, la modélisation de ce type d'inférences est actuellement étudiée, afin de construire un module d'adaptation. Dans ce but, les principes du raisonnements à partir de cas sont utilisés et, en particuliers, les notions de chemins de similarité et de reformulations (Melis *et al.*, 1998).

Des comptes-rendus du CCP ont été analysés par une psycho-ergonome (Sauvagnac, 2000). Une acquisition de connaissances d'adaptation utilisant ces comptes-rendus a été menée, à l'aide de la psycho-ergonome, d'experts en cancérologie et d'informaticiens.

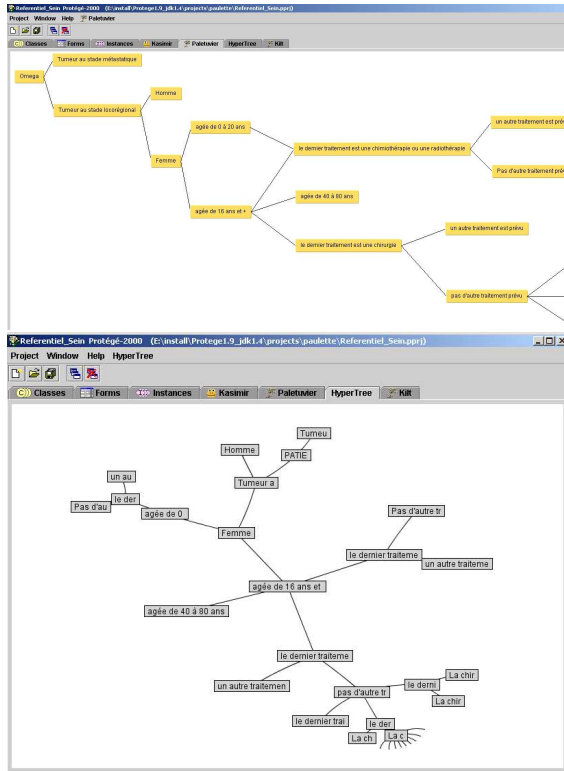


FIG. 3 – Deux modules de visualisation : PALÉTVIER (en haut) et HYPERTREE (en bas). PALÉTVIER est utile, en particulier, pour la visualisation de hiérarchies avec héritage multiple sans duplication de nœuds. HYPERTREE est une API gratuite (Bouthier, 2003) pour la visualisation d'arbres hyperboliques et est un bon moyen de navigation à travers les hiérarchies.

Dans (Lieber *et al.*, 2003), cette acquisition est décrite et, en particulier, plusieurs schémas d'adaptations sont présentés et accompagnés d'une analyse des besoins en représentation des connaissances qu'ils entraînent. Nous allons présenter brièvement trois de ces schémas.

Le premier s'applique quand des données sur le patient sont manquantes. Dans une telle situation, le « critère du minimax (Wald, 1950) » s'applique : la décision (le traitement) est prise sur la base de ses pires conséquences (qu'il faut éviter).

Le deuxième schéma peut être appliqué quand un élément du traitement proposé par le référentiel est contre-indiqué. Cet élément doit alors être substitué par un autre élément de traitement ayant des bénéfices thérapeutiques similaires mais pas les mêmes effets indésirables.

Le troisième schéma s'applique quand une caractéristique numérique d'un patient est proche d'un seuil de décision. Dans ce cas, les deux décisions (celle qui correspond à une valeur inférieure au seuil et celle qui correspond à une valeur supérieure au seuil) sont envisageables, alors que l'application directe du référentiel ne propose qu'une de ces deux décisions. Une implantation d'un moteur de raisonnement s'appuyant sur des seuils de décision flous a été développée pour que deux décisions soient proposées quand la proximité à un seuil est grande (pour plus de détails, voir (d'Aquin *et al.*, 2004)).

6 CONCLUSION ET PERSPECTIVES

Le système KASIMIR est développé dans le cadre du projet KASIMIR dont le but est la gestion des connaissances décisionnelles en cancérologie. Une partie de ce système est destinée à des médecins et est constituée d'une interface utilisateur et de référentiels médicaux représentés dans un formalisme de représentation des connaissances par objets. Une autre partie est destinée aux ingénieurs en charge de la gestion des connaissances et contient plusieurs modules intégrés dans l'architecture de PROTÉGÉ et utilisant le moteur d'inférences de KASIMIR pour l'édition, la visualisation et la maintenance des connaissances. L'exemple de KILT, un module pour comparer deux versions d'un référentiel est détaillé.

Le système KASIMIR ressemble, pour son utilisation dans le cadre du référentiel, à ONCODOC qui implante un guide de bonnes pratiques cliniques pour le cancer du sein (entre autres). La différence majeure entre les deux systèmes se situe au niveau des principes sur lesquels ils s'appuient : ONCODOC s'appuie sur une approche documentaire alors que KASIMIR s'appuie sur une représentation des connaissances et des raisonnements automatiques.

Les recherches informatiques actuelles de KASIMIR suivent deux directions. L'objectif de la première est le développement d'un portail sémantique dédié à la cancérologie, i.e., un serveur Web fondé sur les principes et les technologies du Web sémantique (Fensel *et al.*, 2003) afin de fournir un accès intelligent aux connaissances et aux services utiles pour la cancérologie.

L'autre direction de recherche concerne l'acquisition de connaissances d'adaptation. Trois approches sont en cours d'étude : une approche d'acquisition auprès d'experts, une approche d'acquisition automatique (par examen de la base de cas) et une approche mixte, combinant les deux précédentes.

Remerciements. Les auteurs remercient les autres membres du projet KASIMIR. Ils remercient également les relecteurs de cet article et les prient de les excuser de ne pas avoir pu tenir compte de toutes leurs remarques, notamment pour une question de place.

RÉFÉRENCES

- Baader *et al.*, 2003, rédacteurs F. Baader, D. Calvanese, D. McGuinness, D. Nardi et P. Patel-Schneider. *The Description Logic Handbook*. Cambridge University Press, Cambridge, UK, 2003.
- Bouthier, 2003, C. Bouthier. Hypertree Java Library. <http://hypertree.sf.net/>, 2003.
- d'Aquin *et al.*, 2003, M. d'Aquin, C. Bouthier, S. Brachais, J. Lieber et A. Napoli. Knowledge edition and maintenance tools for a semantic portal in oncology. Rapport de recherche, Loria, 2003.
- d'Aquin *et al.*, 2004, M. d'Aquin, S. Brachais, J. Lieber et A. Napoli. Decision Support and Knowledge Management in Oncology using Hierarchical Classification. Dans *Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP-2004)*, rédacteurs Katherina Kaiser, Silvia Miksch et Samson W. Tu, volume 101 de *Studies in Health Technology and Informatics*, pages 16–30. IOS Press, 2004.
- Evidence-based medicine working-group, 1992, Evidence-based medicine working-group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, 17:268, 1992.
- Evidence-based medicine working-group, 1992, Evidence-based medicine working-group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, 17:268, 1992.
- Fensel *et al.*, 2003, rédacteurs D. Fensel, J. Hendler, H. Lieberman et W. Wahlster. *Spinning the Semantic Web*. The MIT Press, Cambridge, Massachusetts, 2003.
- Lieber *et al.*, 2002, J. Lieber, M. d'Aquin, P. Bey, B. Bresson, O. Croissant, P. Falzon, A. Lesur, J. Lévesque, V. Mollo, A. Napoli, M. Rios et C. Sauvagnac. The Kasimir Project: Knowledge Management in Cancerology. Dans *Proceedings of the 4th International Workshop on Enterprise Networking and Computing in Health Care Industry (HealthComm 2002)*, pages 125–127, 2002.
- Lieber *et al.*, 2003, J. Lieber, M. d'Aquin, P. Bey, A. Napoli, M. Rios et C. Sauvagnac. Acquisition of adaptation knowledge for breast cancer treatment decision support. Dans *9th Conference on Artificial Intelligence in Medicine in Europe 2003 - AIME 2003*, Protaras, Chypre, rédacteurs Michel Dojat, Elpida Keravnou et Pedro Barahona, Lecture Notes in Artificial Intelligence 2780, pages 304–313, Oct 2003.
- Lieber et Napoli, 1998, J. Lieber et A. Napoli. Correct and Complete Retrieval for Case-Based Problem-Solving. Dans *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*, Brighton, United Kingdom, rédacteur H. Prade, pages 68–72, 1998.
- Melis *et al.*, 1998, E. Melis, J. Lieber et A. Napoli. Reformulation in Case-Based Reasoning. Dans *Fourth European Workshop on Case-Based Reasoning, EWCBR-98*, rédacteurs B. Smyth et P. Cunningham, Lecture Notes in Artificial Intelligence 1488, pages 172–183. Springer, 1998.
- Napoli *et al.*, 1994, A. Napoli, C. Laurenço et R. Ducournau. An object-based representation system for organic synthesis planning. *International Journal of Human-Computer Studies*, 41(1/2):5–32, 1994.
- Noy *et al.*, 2000, N. Noy, R. Fergerson et M. Musen. The knowledge model of Protégé-2000: Combining interoperability and flexibility. Dans *Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2000*, October 2000.
- Riesbeck et Schank, 1989, C. K. Riesbeck et R. C. Schank. *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1989.
- Rios *et al.*, 2003, M. Rios, E. Desandes, B. Bresson, I. Klein, A. Lesur, F. Boisson, V. Demange et P. Bey. Référentiels de bonnes pratiques cancérologiques : étude comparative de trois supports d'aide à la proposition thérapeutique pour les cancers du sein et de la prostate en Lorraine. *Bulletin du cancer*, 90(4), 2003.
- Sauvagnac *et al.*, 1999, C. Sauvagnac, J. Stines, A. Lesur, P. Falzon et P. Bey. Application of therapeutic protocols: a tool to manage medical knowledge. Dans *Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM-99)*, rédacteurs W. Horn, Y. Shahar, G. Lindberg, S. Andreasen et J. Wyatt, volume 1620 de *LNAI*, Berlin, 1999. Springer.
- Sauvagnac, 2000, C. Sauvagnac. *La construction de connaissances par l'utilisation et la conception de procédures. Contribution au cadre théorique des activités métafonctionnelles*. Thèse d'Université, Conservatoire National des Arts et Métiers, 2000.
- Wald, 1950, A. Wald. *Statistical Decision Functions*. Wiley, New York, 1950.